Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research

Arthur B. Kennickell
Senior Economist and Project Director Survey of Consumer Finances
Mail Stop 153, Federal Reserve Board, Washington, DC 20551
Phone: (202) 452-2247
Fax: (202) 452-5295
Email: Arthur.Kennickell@frb.gov
SCF Web Site: http://www.bog.frb.fed.us/pubs/oss/oss2/scfindex.html

Key words: Wealth measurement, Survey of Consumer Finances

Surveys are the large scientific measurement devices of the social sciences. Each set of observations rests heavily on theories—statistical theories about sampling, missing data, inference, and other such formal issues, cognitive theories about how people perceive the data collection process and respond to it, behavioral theories that deal with the different incentives that operate on all parties in the data collection, and even social theories that address the ways that we interpret the process and its outcomes. Each aspect of these theories has a role in characterizing the "results" of the cumulation of processes that make up a survey.

Despite what is usually a high degree of commonality at the level of basic theory, most surveys differ in details that are often highly important in interpreting the data they collect and in making comparisons across surveys. In household surveys of economic behavior, there are three basic problems: first, respondents often have a poor understanding (or sometimes a different one from that of analysts) of many key economic concepts; second, they usually consider their economic information to be private; third, some items and types of behavior are relatively concentrated in a relatively small group of people. The need to respond to these problem throws particular strands of data collection issues into relief. This paper gives a systematic overview of the Survey of Consumer Finances (SCF), which is designed to collect detailed information on households' finances in a broad sense, with a focus on the research that has been done to support the design, operation, and analysis of the survey.[1]

The first section of the paper describes the content of the survey and briefly reviews the history of the SCF. The next section focuses on the structure of the survey since the major technical revisions in 1989; the first part of this section discusses the sample design and factors related to the construction of the analysis weights, the second part examines issues related to the actual data: the questionnaire design, the collection of the data, response issues, data processing, and data quality, and the third part addresses the disclosure limitation issues in the creation of a public version of the dataset. The final section of the paper summarizes the directions where further research to support the survey is needed.

---

[1]See Kennickell, Starr-McCluer, and Surette (2000) for a recent summary of descriptive results from the survey.

I. Background on the SCF

A. Content of the survey

The SCF is intended to collect detailed information on all household assets (including residences, other real estate, businesses, all types of financial assets, pensions, and other assets) and liabilities (including mortgages, installment loans, credit card debt, pension loans, and other debts) along with auxiliary information useful for analyzing the wealth data (income, demographics, marital history, employment history, attitudes, etc.).[2] To that end, the survey uses a highly structured instrument to elicit the sort of information that carefully qualifies the types of assets and liabilities owned by the survey households, as well as for related characteristics of those items. For example, for each actively managed personal businesses, the survey asks for the nature of the business, how the business was obtained, the number of employees, the legal form of ownership, the nature of the household's loans to and from the business, the value of the business, the tax basis, the gross revenue, and the net income. For most loans, respondents are asked the purpose of the loan, the size of the payments, the term of the loan, the interest rate, and the institution to which the debt it owed. Over the course of the interview, answers to the asset and liability questions are used to build a roster of the financial institutions that provide the financial services used by the household, and a set of questions is asked about those institutions.

A significant cost of breadth and detail of the questions is the length of the interview. In the 1998 SCF, the most recent of the series, the median interview length was about 75 minutes, and some very complex interviews took more than three hours. The introduction of computer assisted personal interviewing (CAPI) in the 1995 survey has shortened this time somewhat compared with earlier surveys.

Because holdings of many types of assets and liabilities are highly concentrated in a relatively small fraction of the population, a typical equal-probability sample would yield too few cases for analysis for many analytical purposes unless the sample size was quite large. As discussed in detail later in this paper, the SCF employs a special sample design to oversample wealthy families.

---

[2]See the "Codebook for the 1998 Survey of Consumer Finances" (Kennickell, 2000a) for a detailed list of variables and corresponding questions.

B. A brief history of the SCF

The intellectual foundation of the SCF began with the 1962 Survey of Financial Characteristics of Consumers (SFCC), which was directed by Dorothy Projector at the Federal Reserve Board.[3] The survey was specifically focused on measuring the components of net worth. The instrument design was unique to that point in survey history in the depth to which it probed household finances; some respondents were even asked to list their individual holdings of stocks. However, a more fundamental insight was the idea of oversampling wealthy households. The SFCC used 1960 Census data along with statistical records derived from tax returns by the Statistics of Income Division (SOI) of the IRS to classify households into strata defined in terms of income. The stated rationale for this sample design was to increase the efficiency of estimates of highly concentrated assets. However, it is quite clear even in those days of generally higher survey response rates, that households in the more affluent strata were less likely to participate.[4] Given this differentially higher nonresponse, many of the survey estimates would have been biased had it not been possible to make some correction along the wealth dimension using the original design information. This insight, which was apparently unacknowledged at that time, is a critical factor in the operation of the later SCF.

Data quality reached what is in some ways an historical high point in the SFCC. Skilled technical people at the Federal Reserve Board actually looked up market values for the securities respondents enumerated and took other such extraordinary steps to ensure the highest information quality. In addition, a reinterview—the Survey of Changes in Family Finances (SCFF)—was conducted in 1964 (to collect 1963 data), and respondents were asked to "correct" their earlier data at that time. After years of data processing, much of which was done by hand, a tape of the survey data was released to the public in what was one of the pioneering examples of public use datasets.

---

[3]Data were collected in 1963, but the values requested were generally as of December 31, 1962. See Projector and Weiss (1966) for descriptive analysis and more details on the design and execution of the survey.

[4]According to the surviving verbal history of the survey, President Kennedy made some telephone calls to convert refusals. Clearly, privacy and confidentiality issues were less of a problem at that time.

Some evidence survives of another iteration of the survey conducted in 1965 (to collect 1964 data). Unfortunately, there was never funding to keypunch the data, and ultimately all the paper questionnaires were lost.

The need for information as an input to economic policymaking and related research is strong, and generally the key inputs are current data and time series of historical data. However, in the short run, the incentives to build time series are weak if meaningful results are not available until years after the time of data collection. The incentives are particularly weak when the time series are generated by such an expensive measurement device as a survey. Faster data processing facilities might have allowed the SFCC to survive by providing more current data while waiting for the series to build up. In the end, the survey was not renewed, and the next comparable wealth survey was not until 20 years later.

Meanwhile, some survey research on wealth-related issues continued elsewhere. Starting after World War II, surveys called the "Survey of Consumer Finances" were run at the Survey Research Center (SRC) at the University of Michigan, principally under the direction of George Katona.[5] These surveys began with a focus on purchase intentions, with credit information collected (and some asset and income data) to serve in the interpretation of the purchase intentions. Eventually, sufficient policy interest in survey data had revived that the Federal Reserve provided funding for the 1977 Survey of Consumer Credit (SCC) with Thomas Durkin as the project director and Richard Curtin as the project manager at SRC.[6] That survey synthesized much that had been useful in the earlier Michigan surveys with some significant extensions to provide a picture of household credit use that was unprecedented to that time. Although little was done to collect wealth data, the perspective on data collection fed directly into the framework for collecting consumer credit data in the later SCF.

---

[5]For general information on the 1967 survey and references to the earlier surveys, see Katona et al. (1968).

[6]See Durkin and Elliehausen (1978) for descriptive data and details on the survey methodology. The Office of the Comptroller of the Currency and the Federal Deposit Insurance Corporation also sponsored this survey.

The rebirth of wealth measurement at the Federal Reserve with the 1983 SCF sprang from a complicated confluence of events.[7] The Federal Reserve wanted to repeat the SCC to and collect some additional data on deposits, both of which were also of interest to the Office of the Comptroller of the Currency and the Federal Deposit Insurance Corporation. The Department of Health and Human Services (DHHS) wanted to collect information on pension rights, and there was latent interest in other agencies that was ultimately linked. In the background, three people operated in critical moments and in complicated ways to bring the modern SCF to birth in 1983: Glenn Canner and Barbara Lowery at the Federal Reserve and Fritz Scheuren then at SOI. Robert Avery became the director of the project, and he led a number of other very talented people in the design of the questionnaire for the 1983 SCF. Thomas Gustafson at DHHS was instrumental not only in later design issues, but also in arranging the largest share of the funding, which allowed the project to move beyond the design stage. Again, Richard Curtin served as the SRC project manager, and he made substantial contributions to the design. The survey was fielded in June of 1983.[8]

Beyond the questionnaire, the key to making the 1983 SCF credible as a wealth survey was the dual-frame sample similar to that used in the SFCC. One part of the sample was drawn using an area-probability design by Steven Heeringa at SRC. The other part was a list sample selected by Tom Petska at SOI under the guidance of Fritz Scheuren on the basis of income characteristics. Permission to use SOI data for this second sample was difficult to obtain, and there were severe constraints placed on how the sample data could be used. Among the most significant limitations was that the list sample respondent could only be contacted once, and they could only be interviewed if they returned a postcard included with the mailing; the result was a "volunteer" sample of only about 10 percent of the original mailing. Despite stringent

---

[7]For an overview of the SCF variables and a more detailed acknowledgment of the key players in the survey, see Avery and Elliehausen (1983). For an evaluation of the survey data, see Avery, Elliehausen and Kennickell (1988). The 1983 survey also had a companion piece of data collection in the Survey of Pension Providers, which was based on summary plan descriptions for the pension plans offered by employers of SCF respondents who provided information to contact their employers; see Curtin (1985).

[8]The survey had originally been intended to be conducted in 1982, but a longer-than-expected design phase and other factors delayed the start.

precautions at the time to protect the privacy of respondents, there was still notable political controversy about the operation, and one key player nearly lost his job as a result. Nonetheless, the survey was completed, and it was widely viewed as highly successful both as a source of current information and as an input to longer-term research.

About three years later–an interval determined largely by what could feasibly be accomplished in processing the data and providing preliminary analysis of the 1983 SCF–planning began for a reinterview of the 1983 sample.[9] However, the interest in additional credit data and some change information for policy analysis at other agencies was only sufficient to generate funds to pay for a telephone interview. To accommodate the time constraints of a telephone interview, the wealth questions were dramatically abbreviated to a form roughly similar to the current wealth module on the Panel Study on Income Dynamics. Although the 1986 SCF was useful for a number of purposes, it also failed in significant ways, most of which were a function of the very limited data collected. One positive result of the failures was a commitment at the Federal Reserve not to do another survey unless it could provide information more like what had been obtained in the 1983 survey.

The 1989 SCF marked a major turning point in the evolution of the survey in two important ways. First, a major questionnaire revision was undertaken to address problems that had become clear in analyzing the data from the 1983 survey and to extend the survey to areas where additional related information could be collected. Second, it was the start of a consistent application of methodology that continues through the most recent survey in 1998. This methodology is described later in this paper.

As in the past, sponsorship of the 1989 survey was complicated. Although the Federal Reserve made a substantial contribution to the funding, major funding was also provided by the National Institute on Aging (through a grant for which Thomas Juster at SRC was principal investigator), and smaller amounts were provided by other agencies. In large part, the structure of the funding determined that the survey would have both cross-sectional and panel

---

[9]See Avery and Kennickell (1988) for an overview of the survey and a description of the data collected. Avery, Elliehausen and Kennickell (1988) provides descriptive results from the survey.

dimensions.[10] The cross-section has proven to be the most useful product by far. Processing of the panel component was difficult, expensive, and time-consuming, and owing both to sample attrition and reporting problems, the data quality appeared to be somewhat questionable.[11] Since 1989, the possibility of an SCF panel has recurred many times, but the increasing difficulty of convincing respondents to participate in even a single SCF has made the possibility of a successful reinterview more unlikely than in 1983. This author's hope is that further research in collecting retrospective data will lead to means of measuring what we *need* to know (as opposed to what we may in the ideal *want* to know) to advance our understanding of the dynamics of consumer finances. Since 1989, the survey has been conducted as a cross-section every three years, and it has been sponsored by the Federal Reserve Board in cooperation with SOI. Beginning with the 1992 survey, management of the survey data collection moved from SRC to the National Opinion Research Center (NORC) at the University of Chicago.[12]

II. SCF Methodology from 1989 Forward

A. Sample issues

1. Sample design

The SCF is required to provide estimates of items that are broadly held across the population–such as checking accounts and credit cards–as well as assets and liabilities that are held by smaller groups–most often, wealthier groups–or are relatively concentrated by value among smaller groups. Direct ownership of bonds, for example, is relatively rare overall but more common among wealthy households, and net worth in general is highly

---

[10]See Heeringa, Connor and Woodburn (1994) for a detailed description of this design.

[11]See Kennickell and Shack-Marquez (1992) for a description of the cross-sectional data, and see Kennickell and Starr-McCluer (1997a, 1997b) for a discussion of the panel data and some results.

[12]It is possible that the change in survey vendors between 1989 and 1992 altered the data collection process in a way that affected the distribution of outcomes. However, the project staff at the Federal Reserve Board have always maintained a very strong involvement in the design and execution of interviewer training, and they have also exercised intensive oversight of field operations. Moreover, data processing beyond the data entry stage and initial coding stages has always been largely conducted by the Federal Reserve staff.   Since these are the most critical points at which the data might be affected, the "house effect" in looking at changes between the 1989 and 1992 SCFs is likely to be quite small.

concentrated–about 34 percent of the total was held by the wealthiest one percent of the population in the 1998 SCF (Kennickell, 2000b). To support estimates of a variety of financial characteristics, the survey employs a dual-frame sample design.

A national area-probability (AP) sample provides good coverage of widely spread characteristics.[13] The sample selects household units with equal probability through a multistage selection procedure. At the first stage, a group of major metropolitan areas (19 such areas in the sample used for the 1995 and 1998 SCF) are selected as primary sampling units (PSUs) with probability one; for the remainder of the country, counties and metropolitan statistical areas are stratified by a variety of characteristics, and PSUs are selected proportional to their population. At the second stage, smaller areas are selected within PSUs to serve as the basis for the ultimate samples. Finally, at the time of the execution of the SCF, housing units within selected areas are enumerated, and a set of ultimate observations is drawn.

Although an AP sample is an efficient means of generating a nationally-representative sample, by itself it has two important shortcomings for the SCF. First, given the concentration of ownership of many assets, it is very unlikely without an enormous sample size, that an AP sample would yield sufficient observations for the analysis of many types of financial behavior. Second, there is ample evidence (discussed below) that there is nonrandom nonresponse in the SCF, and that wealth (or variables correlated with it) is a key factor in explaining that nonresponse.

To deal with these two problems, the SCF also employs a list sample. This sample is developed from statistical records derived from tax returns under an agreement with SOI.[14] To

---

[13]See Tourangeau *et al.* (1993) for a complete description of the design and selection of the area-probability frame used for the 1995 and 1998 SCF. This frame was based primarily on 1990 Census data. The 1989 and 1992 surveys used an earlier frame constructed using 1980 Census data; the selection of this sample was quite similar to the procedure used in the later surveys.

[14]See Wilson and Smith (1983) and Internal Revenue Service (1992) for a description of the SOI file. The file used for each survey largely contains data from tax returns filed for the tax year two years before the year the survey takes place. This file is based on a sample of tax returns, rather than the universe of returns. However, because the sampling rate is so high in income regions that turn out to be most important for the SCF, this limitation is generally ignored in using the file for sampling for the SCF. See Kennickell (1998c) for a detailed

protect the privacy of taxpayers, this agreement places strong restrictions on the use of the information provided by SOI as well as on the treatment of the ultimate survey data.

The income information in the SOI dataset is used to compute a "wealth index," which corresponds roughly to a capitalization of income flows, and the observations in the SOI file are stratified using this index. Observations are sampled with progressively higher rates from strata corresponding to higher levels of the index. In the 1989 and 1992 SCF, the wealth index was computed as the sum of the product of a number of income types and the inverse of the average rate of return associated with that type of income.[15] After the 1992 survey was completed, a complicated negotiation allowed the linking of a measure of net worth and a very limited set of other variables in the SCF with a selection of information in the SOI file for following two reasons only: first, evaluating the performance of the wealth index, and second, for estimating an alternative model-based wealth index. Access to the linked information was restricted to only this author.[16] For the 1995 and 1998 SCF, the final wealth index is a blending of a version of the original index and one generated by a model estimated using data from the previous survey. The blended form was used for two reasons. First, based on an ex post evaluation, there appears to be a gain in statistical efficiency from using the two estimates. More importantly, the blending hedges against the possibility of nonstationarity in the estimated model, where the model coefficients are implicitly functions of the structure of rates of return and regulations and other institutional constraints in place at the time of the survey on which the model is estimated.

There are several potentially important problems with the list sample.[17] The elements in the SOI file are taxpayers, but the target population in the SCF is households. Where a household consists of either a married couple filing a joint tax return or a single person filing a return, possibly along with dependents, the information in the SOI file may represent the

---

[14](...continued)
description of the selection of the 1998 list sample.

[15]For example, if the average rate of return on interest-bearing assets was 5 percent, then interest income would be inflated by a factor of 20 to estimate the stock of such assets. See Heeringa, Conner and Woodburn (1994) for details.

[16]See Kennickell and Woodburn (1999), Kennickell (1998a), and Kennickell (1998b) for more details.

[17]See Kennickell and McManus (1993) and Frankel and Kennickell (1995) for a more detailed evaluation of problems with the list sample.

household well. However, many households have more complex situations. In the case where a married couple files tax returns separately, adjustments can be made at the sample selection stage. However, for couples living together as unmarried partners and for other households with multiple adults, the approximation error in using a single taxfiler's income may be large. Fortunately, investigation of the filing patterns of households suggests that married couples in relatively wealthy households are more likely to file joint returns and that other returns for such households are not likely to alter the picture of their finances much. At the other end of the economic spectrum, the problem in the list sample is likely to be larger. In addition, the proportion of households that file no tax return is not negligible among poor households. These findings feed directly into key choices in the weighting design for the survey that are intended to minimize the effect of such problems.

Examination of net worth as a function of the wealth index (figure 1) suggests that wealth and the index are clearly closely related, and for most cases in the upper end of the wealth distribution, the classification is better than at lower levels.[18] Nonetheless, there is still very substantial scope for "misclassification." In addition to errors introduced by mismeasurement of the appropriate unit in the SOI data, there are two key income-driven problems.

First, some key assets may leave no trace in a given year of a filer's tax data. For example, a principal residence does not ordinarily generate income that would appear on a tax return. However, the data suggest that the problem of omitted asset income is not likely to move the ordering implied by the wealth index substantially away from that implied by actual wealth.

Second, a related problem is that rates of return may vary widely across individuals for assets that generate income at least occasionally. For example, a business may have great market value but may generate negligible income at some stages of its development. When sold, such assets that generate minimal return may appear as large capital gains. The survey data suggest that rate of return variability can be quite large, and it probably introduces substantial noise into the wealth index as a proxy for wealth. Unfortunately, there is very little that can be done to address this problem directly with only a single cross-section of SOI data. Access to

---

[18]For ease of interpretation, the horizontal scale of the plot is shown in terms of percentiles of the unweighted distribution of the wealth index over the sample participants.

multiple years of SOI data could help to smooth out some rate of return fluctuations. Movement toward a design based on multiple years of data is a key objective for the 2001 SCF.

Figure 1: Distribution of wealth index by unweighted deciles of net worth, 1998 SCF.

Percentile of distribution of wealth index
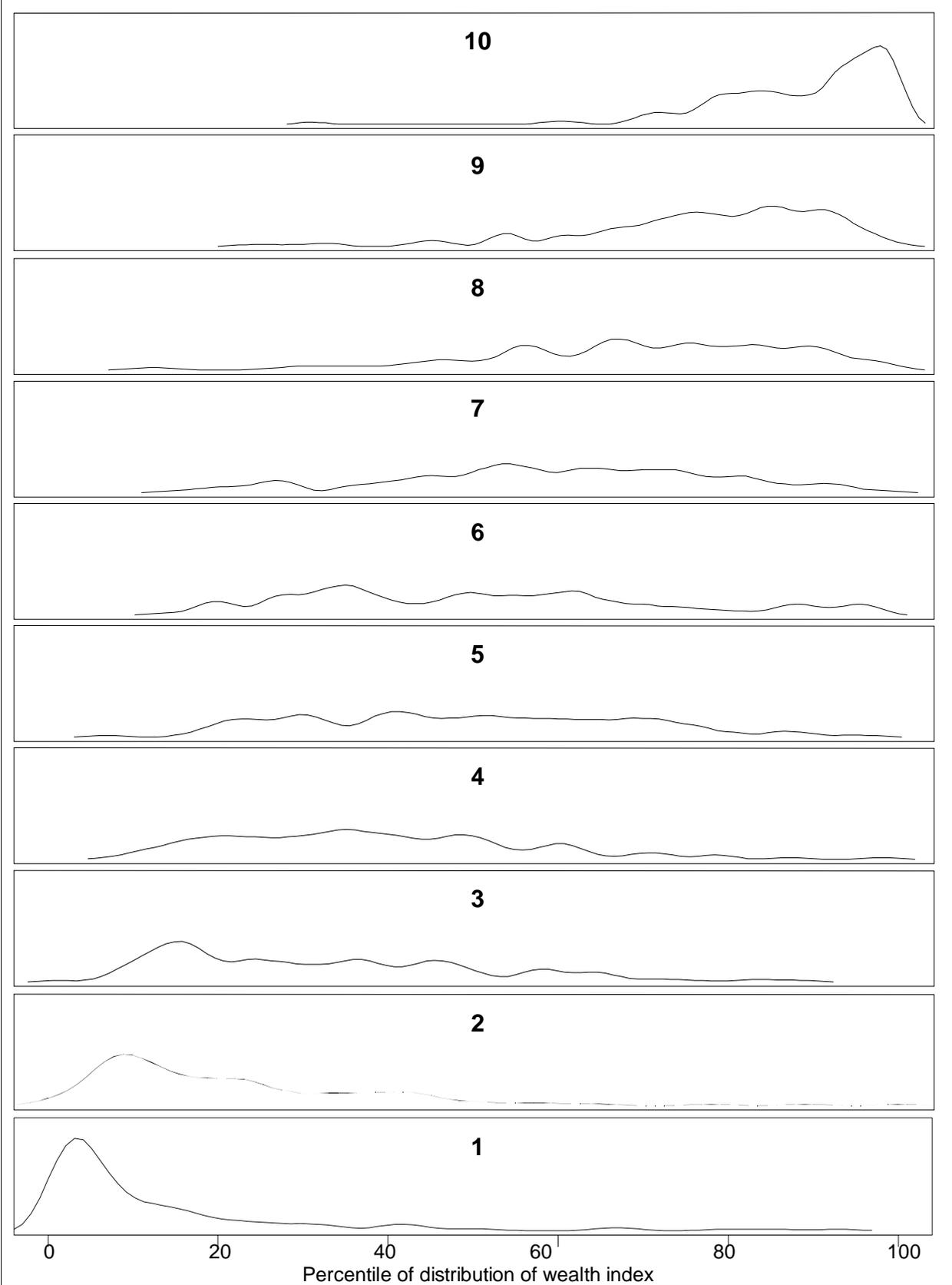
The final pressing problem in the design and use of the list sample to be raised here is the geographic distribution of that sample. For reasons related to cost control on the survey, the geographic distribution of the list sample is constrained to that of the area-probability sample. However, there are clear differences between the population density of wealthy households and that of all households. Where the major population centers of wealthy households overlap with the PSUs of the AP sample, there is no problem. However, there are omitted population centers. The most notable omission before 1995 was Palm Beach, Florida; this locality has been a PSU in the AP sample used for the SCF since that time. In most other areas (metropolitan statistical areas and county units defined comparably to those used in the design of the AP sample), the number of omitted wealthy cases available for sampling is fairly small, and it is not clear that their inclusion would have important effects beyond what is already achieved by post-stratification at the weighting stage.

2. Unit nonresponse

No matter how much care one puts into a sample design, there is usually little that one can do directly to ensure that the final set of sample participants is an unbiased subsample of the full sample. As one might expect given the sensitive nature of the subject matter of the SCF, nonresponse in the survey is a very important problem, it has clearly nonrandom components, and it has shown signs of becoming worse at least since 1992 (table 1). In the AP sample, the response rate was less than 70 percent of eligible cases in every year since 1989; as shown in the table, the response rate for this sample is inversely related to the population size of the PSUs. The overall list sample rates are only about 30 percent, and even this low figure masks substantially lower cooperation rates in the higher wealth index strata.[19]

---

[19]As a part of the agreement with SOI that allows the use of their data for sampling, list sample cases are given an opportunity to decline participation in the survey before they are contacted by an interviewer. Starting with the 1989 survey, the list sample cases were sent material explaining the survey and a postcard to be returned if they did not wish to participate. The proportion of the list sample returning the postcard was 38.1% in 1989, 23.4% in 1992, 21.2% in 1995, and 13.5% in 1998. The decline in the fraction of returned postcards suggests that the rise in the level of junk mail over this period led an increasing fraction of respondents to discard the informational packets without reading them. All cases returning the postcard are treated as eligible respondents for purposes of calculating the response rate.

Table 1: Number of completed cases and response rates by sample type, 1989, 1992, 1995 and 1998 SCF.

	Survey year			
	1989	1992	1995	1998
Number of completed cases:				
Total	3,803	3,906	4,299	4,309
AP sample	2,277	2,456	2,780	2,813
List sample	866	1,450	1,519	1,496
Response rates:				
AP sample[*]	64.3	68.0	66.3	65.9
Self-representing PSUs	NA	61.8	58.9	62.3
Other MSA	NA	67.4	66.6	66.6
Other areas	NA	75.7	77.6	70.3
List sample[†]	29.5	31.3	30.4	28.6
Stratum 1	40.5	42.8	45.3	41.3
Stratum 2	38.4	41.4	39.5	39.2
Stratum 3	36.2	37.4	35.5	36.2
Stratum 4	34.2	34.7	35.0	35.8
Stratum 5	25.8	31.4	30.4	30.4
Stratum 6	17.5	26.0	23.9	23.9
Stratum 7[‡]	NA	14.4	12.8	8.3

[*] The overall response rate for the 1989 AP sample reported applies to the 1989 new cross-section cases only. In addition, the decomposition of the response rate by PSU type is not available for the 1989 survey.
[†] The definitions of the strata differ across the survey years, but in each year, strata 1 and 2 correspond roughly to the types of cases in the AP sample, and the higher numbered strata are progressively more wealthy.
[‡] Only six strata were created in 1989.

Several key factors contribute to unit nonresponse: the resistance respondents feel to participating in a survey, the ability of interviewers to interact with respondents and communicate reasons for participating, the willingness of interviewers to attempt cases, and the overall allocation of field resources. Obviously, the best way to deal with the nonresponse problem is to eliminate it. The development of strategies to minimize nonresponse has always been a key element of research for the SCF.

To this end, interviewer training is a critical part of the survey cycle. During the training, every possible effort is made to ensure that interviewers understand the motivation for the survey and that they can communicate a compelling argument for why a respondent should want to

participate. In addition, they are given training in how to respond to questions, such as those related to confidentiality, that have often been raised by past respondents. The training for the 2001 survey proposes to use professional actors to impersonate a variety of respondents in order to allow interviewers to work out realistic strategies within a supportive training environment. Over time, an array of materials—brochures, letters, etc.—has been developed to aid interviewers in gaining the cooperation of respondents. A toll-free number and, more recently, a web site offer respondents a means of verifying the legitimacy of the survey. The 2001 survey also proposes to develop a video to use to gain permission from property managers to enter a locked building or gated community.

An important route to improving response on the survey is a better understanding of the processes that generate lead to nonresponse. At least for the list sample, there is a substantial body of information from the original frame about the entire sample. In addition to the economic data in the file, it also contains some demographic information. Beginning with Kennickell and McManus (1993), this source has been exploited in hopes of understanding the nonresponse process more clearly and making better statistical corrections for it through weighting adjustments.[20] Several points are quite clear in the data. First, nonresponse is highly correlated with the wealth index, as can be seen directly from the response rates by stratum. Second, the most powerful correlate of nonresponse in a complex model using a variety of frame data is a measure of financial income, defined as the sum of interest and dividend incomes. Third, there are a variety of age and geographic effects.[21] For the AP sample, other than the population data from the frame, the only other information available consistently since 1989 that can be used to characterize the nonresponse patterns is the Current Population Survey (CPS). Comparisons of the age distribution and the home ownership rate, factors that are measured with sufficient similarity in both surveys, show systematic differences. Particularly striking is the

---

[20]This is one of the set of narrowly circumscribed uses permitted of these data.

[21]These findings summarize nonresponse for this sample overall. The research also suggests that the process that underlies nonresponse by virtue of having returned the postcard that allows list sample members to opt out of the survey is different from that underlying nonresponse conditional on not returning the postcard. However, the models explaining overall nonresponse and nonresponse conditioned on not returning the postcard differ mostly in terms of the magnitudes of their coefficients rather than in terms of their signs.

under-representation of homeowners in the unweighted SCF AP sample. These results for the list and AP samples are a key underpinning of the weighting design discussed below. Other research has deepened the understanding of the mechanism of nonresponse, both for the AP and list sample cases.

The interactions of the respondents and the interviewers allow the interviewers to apply their training and to draw on the support materials. As Groves and Couper (1996) have argued, for the negotiation with the respondent to be effective, the interviewer must listen to the respondent and tailor the information presented to suit the situation. Respondents differ in their understanding of the survey process and in their willingness to share private information. Interviewers differ in the depth of their belief in the messages delivered in training and in their ability to build effective arguments to deal with points raised by respondents. Evidence from the 1995 SCF (Kennickell, 1999b) suggests that there are identifiable characteristics of respondents and interviewers that make a noninterview more or less likely. Some results, such as the negative correlations between response rates and the average commuting time for people in the Census tract and of average house value, have an economic interpretation as signals of the value of time. Other factors, such as the positive association of the average level of education in the Census tract with response, have a more natural interpretation as an amplifying factor in a model that explains the exchange of information in negotiating agreement to do an interview. Although the results of the research are highly suggestive and they summarize formally what was previously just anecdotal evidence, there is no obvious direct implication for changes in interviewer training or in the interviewer's role beyond what is done already. However, the importance of understanding the function of the respondent-interviewer interaction argues strongly for additional attempts at modeling this behavior as well as for additional measurements.

More recently, research on the survey (Kennickell, 1999c) has looked at cases that were interviewed "early" or "late" in the survey field period for the 1998 SCF and compared those cases to the nonrespondent population.[22] If the later cases were more like nonrespondents than

---

[22]Several measures of "lateness" were considered: a strictly time-based measure, one based on the number of attempts made to secure an interview, one based on the number of

they were like other participants, this situation could have implications for a range of issues, particularly the value of persistence in targeting field resources and the possibility of differentially weighting later observations. The work suggests, perhaps surprisingly, that the differences between early and late cases overall are not sufficiently clear to be useful either in understanding the dynamics of interviewing or in making finer adjustments to the analysis weights to compensate for nonresponse. For the AP sample, the data indicate that the nonrespondent cases tend to be generally wealthier and better educated than any of the participants, but in some dimensions they tend to be somewhat more like the late cases. For the list sample—in which there were explicit minimum quotas by stratum—any patterns in the data are obscured by complex layers of management decisions about which cases to pursue, and by what may be an interviewer aversion to cases in that sample.[23]

The software that is now used to track interviewers' field efforts makes it possible to observe much more clearly the application of effort to the survey cases. However, so little is generally known about individual respondents in advance that it is not possible to monitor whether effort is being applied appropriately to all cases.[24] In the recent surveys, interviewers have been asked to complete a neighborhood and respondent observation form for every case. This information has been quite useful in the attempts to analyze the patterns of nonresponse.

---

[22](...continued)
contacts, and a hybrid measure of the other three.

[23]In related work, Kennickell (2000c) uses SCF data to argue that informational asymmetries between interviewers and survey administrators in the presence of economic incentives that are largely independent of the characteristics of individual cases act to alter the distribution of effort applied to cases, and consequently the distribution of the characteristics of the ultimate survey participants.

[24]A deeper point is that the optimal level of effort is often not well defined. The following examples illustrate this point. If cases differ in the amount of persuasion it takes to obtain an interview, then an equal application of effort will under-represent these more skeptical people. Alternatively, if effort were targeted in a way that was proportional to the level of difficulty, then the resulting data would be contaminated by a treatment effect if behavior within an interview is differentially and adversely affected by the initial effort. Another approach might be to subject all cases to a more intense initial experience; however, because people differ in their responses to pressure, this approach could well have the effect of changing the distribution of participants to favor those who respond to pressure, and inducing a treatment effect in participants who would have been less affected by an approach more tailored to their situation. Whether directly or indirectly, every survey must decide what are its critical points of balance.

However, it may be that this information, combined with neighborhood characteristics obtained from the Census, could be useful in targeting resources to achieve a better balance of participants, at least at the level of the factors used in the post-stratification applied in the weighting adjustments for nonresponse. The 2001 SCF is planned to include additional data capture with the goal of developing at least a framework for improved control of effort applied to cases in the field. Also planned for that survey is an observational study of the sample neighborhoods in two PSUs to search for factors connected with nonresponse that may be missed in using indirect or more aggregated data.

Given the experience of the 2000 Census at the time of writing this paper, it appears that the prospects for further declines in survey participation are threatening unless new means are found to motivate respondents. In the SCF, increasing efforts have been devoted to creating better materials for interviewers to use in persuading people to participate, to providing more realistic training to interviewers about how to counter the arguments of resistant respondents, and to increasing the resources devoted to nonresponse conversion. Still, it appears likely that response rates will fall further even as costs continue to escalate. The nonresponse problem is common to most surveys, and it is possible that progress may be made by developing a more collective strategy for countering the largely negative information that people are given about data collection, privacy, and related issues. One option might be for a credible independent body, such as the American Statistical Association or the National Academy of Sciences, to develop a continuing educational campaign to clarify to individuals the importance of knowing about them if all types of people are to be represented in a meaningful way in policy debates and in the longer term research that ultimately feeds into public policy. Continuing scientific support for data collection would be unlikely to reach the intensity of the Census efforts, but it seems likely that such pressure would have diminishing returns if it were sustained for a long period anyway. Although a slow building of support may not have the excitement of a large campaign, it offers hope for planting a deeper understanding.

3. Weights

The SCF weights must represent as faithfully as possible the original probability structure of the sample, incorporate adjustments for known dimensions of nonresponse, and optimize the efficiency of the data in estimation of population statistics. Out of a long history of

experimentation and sensitivity testing, a set of weights has been developed for the survey that provides a consistent framework for the surveys beginning with 1989.[25]

A key strategic decision was to combine the AP and list samples through weighting. One could make estimates using the two frames separately, but two factors would complicate such an exercise. First, for reasons related to disclosure limitation, it is not possible to give anyone outside the SCF project group the basic frame information necessary for making dual-frame estimates. Second, each sample is limited in some ways that are offset by the other sample. The AP sample, like most samples, suffers from differential nonresponse that is correlated with wealth, and because it is not reasonable to think that wealth is sufficiently correlated with any of the variables available for constructing post-strata, estimates of tail-sensitive statistics such as the mean of net worth would be biased. The list sample can plausibly be adjusted to some degree to compensate for the differential nonresponse with respect to wealth, but the correspondence of the unit of observation in the underlying sample frame with the intended unit of analysis becomes weaker at the bottom end of the wealth spectrum. Moreover, the list sample does not include anyone who did not file an income tax return, about 13 percent of households according to the 1998 SCF. To exploit the relative strengths of each sample, the two samples are joined using a simple post-stratification scheme described below.

The analysis weights are constructed in three stages. First, each sample is adjusted separately using all the useful information that can be brought to bear in creating post-strata. For the AP sample, the probability weights are post-stratified to totals for geographic areas, age group, and home ownership status. For the list sample, the probability weights are adjusted using national estimates of the stratum totals, information on location and a measure of financial income. More complex adjustments to the list sample weights, such as using an estimated response propensity score, seem to add little additional information but they induce a much higher level of variability in the weights. At the second stage, each household that did not file tax returns is given its adjusted AP weight, and all other observations are classified according the gross assets of the household. The weights of each sample are rescaled within these groups to

[25]The evolution of this work is summarized in Heeringa, Conner and Woodburn (1994), Kennickell and Woodburn (1992), Kennickell, McManus and Woodburn (1996), Kennickell and Woodburn (1999), and Kennickell (1999d)

reflect the number of cases contributed by each sample. The totals in the top groups are adjusted to sum to the estimates from the list sample alone, and remaining cases are rescaled to equal a population estimate from the CPS. In the final stage, the weights of the cases in the lower gross asset cells are further adjusted by age, region, and home ownership. When refinements are made to the weight design, they are applied retrospectively to all of the surveys beginning with 1989 in order to maintain as statistically consistent a series as possible.

To enable users to gauge the precision of their estimates, the SCF provides a set of replicate weights to be used in simulating the sampling distribution of estimates. The replicates, which are described in detail in Kennickell and Woodburn (1999), are a structured bootstrap sample of the actual observations. Because of the complex nature of the SCF sample and the levels of adjustments that are imposed on the weights, there is no straightforward traditional alternative to a bootstrap sample. The application of this technique attempts to exploit the key dimensions of variation in the sample. At the first stage of the selection of the AP sample, PSUs outside the self-representing area are drawn in pairs (or sometimes triplets); in self-representing areas, sub-areas are selected in pairs. The bootstrap procedure selects with replacement two (or three, where appropriate) PSUs from each set. The situation for the list sample is a little more complicated. Because the list sample cases are only selected from PSUs that are in the AP sample, the cases in the non-certainty PSUs are selected in parallel with the AP replicates. For list sample cases in the self-representing areas, cases are randomly selected by strata. For each set of AP and list replicates, the full weighting calculation is used to produce a replicate-specific weight. The resulting variance estimates appear to be quite robust to many changes in the selection of the replicates. Nevertheless, because of the importance of reliable variance estimates, research in this area continues.

B. Data issues

1. Instrument design

As noted earlier, the SCF questionnaire covers a broad range of financial and related issues in great detail. In 1998, the SCF interview contained 9,639 variables, though a very large number of those variables could not have been answered by even a respondent with the most complex finances. In a version of that dataset that collapses multiple instances of the same question and reduces the dimension of some auxiliary variables, there are about 3,200 variables.

However, across the entire sample the maximum number of questions answered was 784, and the median number was 434 (see figure 2). Considering only questions with dollar answers, which on average are likely to be more difficult and intrusive, the maximum was 107 responses and the median was 35 (see figure 3). Not surprisingly, the interview takes a long time: the maximum time needed for an interview was over 3½ hours and the median length was 77 minutes (figure 4).

To navigate through such a mass of information and interaction, the questionnaire attempts to frame each type of inquiry carefully. Showcards outlining various types of responses have been developed to support the respondent in answering some difficult questions. Interviewers are given extensive training in using the instrument, but they are not expected to be experts in the subject matter; when a need arises, they have at their disposal a comprehensive glossary of terms as well as question-by-question instructions.

Figure 2: Number of questions answered; all cases, AP cases, and list sample cases; 1998 SCF.
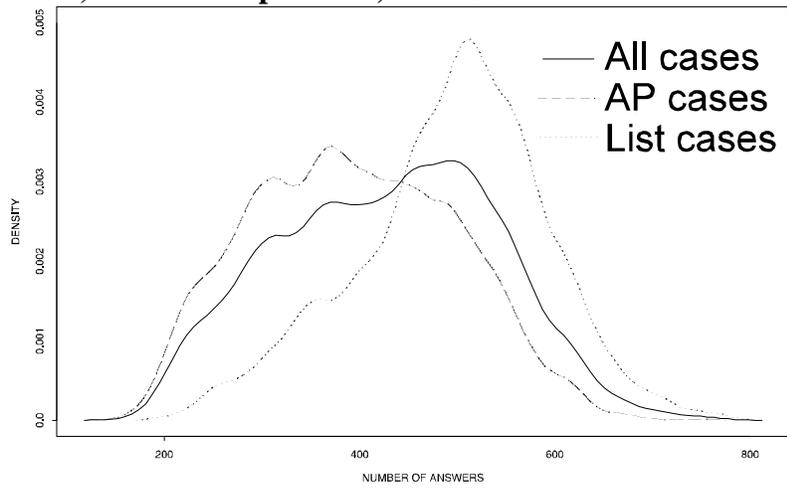

Figure 3: Number of dollar questions answered; all cases, AP cases, and list sample cases; 1998 SCF.
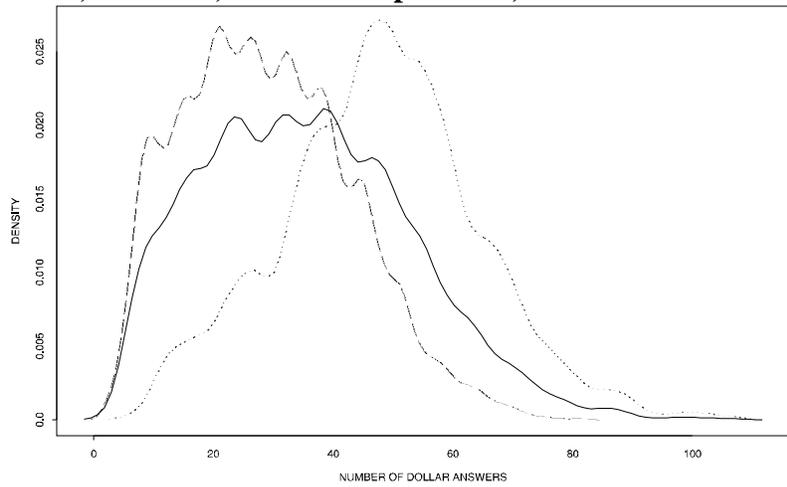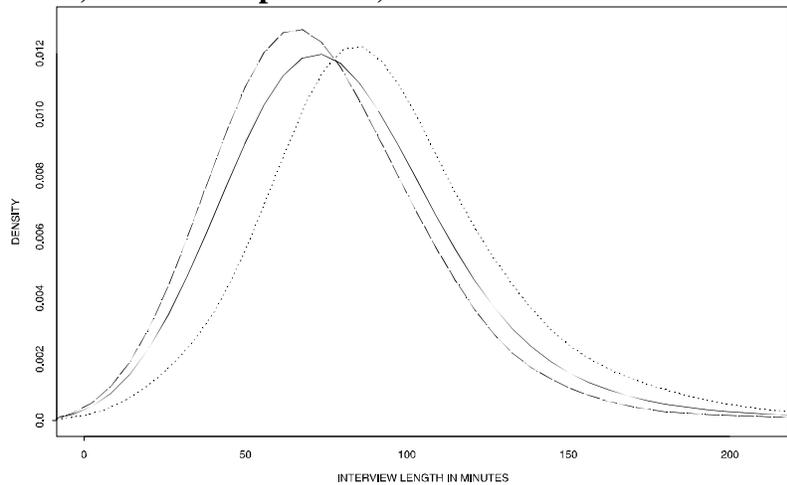

Figure 4: Interview length in minutes; all cases, AP cases, and list sample cases; 1998 SCF.

The SCF migrated from paper questionnaires to CAPI with the 1995 survey. Every possible attempt was made to recreate in the computer program key parts of the range of options available to interviewers in a paper questionnaire. In particular, a means was developed to allow interviewers to break out of repetitive question sequences when the interview was at risk and to request summary data on all remaining items; this technique had been very commonly used by interviewers in the paper framework even though it was not initially even raised as an option in training. CAPI offers many new benefits, of which the following are the most important for the SCF. The data have a uniform logical structure (except in rare cases of computer or software failure). The elimination of the necessity for interviewers to navigate complex skip instructions frees their attention for more important things, such as maintaining the interaction with the respondent and dealing with questions that respondents raise. CAPI also allows immediate access by the interviewer to specialized instructions and explanations. This mode of data collection allows the project staff very rapid access to the data and interviewers' electronic comments, so that computer problems, CAPI programming errors, interviewer errors, and common respondent errors are recognized quickly and a response is returned to the field almost immediately. Finally, it allowed the use of a specialized procedure, which is discussed in more detail below, to automate probing for dollar questions when the respondent is either unable or unwilling to specify a precise amount.

Although the goal of the survey is to maintain as fixed an instrument as possible, inevitably the questionnaire is a continually evolving document. The major 1989 revision incorporated what had been learned in the 1983 SCF and the 1986 reinterview, and it also altered wording of some questions to respond to developments in the financial marketplace since the earlier surveys. Questionnaire revisions have been fairly minor since that time, but for the questions requesting factual data, wording has been changed when analysis of edited data indicates a high error rate or where there have been important market developments. Cognitive interviews and focus groups have helped to explain the nature of some reporting problems and to design remedies.[26] Visual aids—mainly in the form of showcards—are important tools in

---

[26]One example of such work on the SCF is Kennickell, Starr-McCluer and Sundén (1997).

clarifying the framing of some questions, and it is hoped that as electronic data collection progresses, it will be possible to make more routine use of such items with automatic customization to the respondents' circumstances.

It is very clear from analysis of many of the errors that interviewers and respondents have made in reporting data that work needs to move much farther in the visual design of CAPI instruments. Much has been learned about effective layout in paper questionnaires through the extensive work of Don Dillman and others; in essence, this work highlights the fact that needless complications in displaying information are invitations to cognitive error. Unfortunately, much of this knowledge from the world of paper questionnaires had to be temporarily suspended in the movement to the less flexible interface of CAPI. Despite advances in the development of sophisticated interfaces in other parts of computer science, software with both the capacity for collecting complex data structures and the features needed for more controlled screen design is only starting to become available. This opportunity should be exploited to make it easier for interviewers to present information to respondents and to record their answers correctly.

2. Item nonresponse

In general, households that commit to participating in an SCF interview are willing at least to tell whether they have the items the survey asks about (see the second column of table 2). Reporting of the values of such items is more problematic. Some respondents may know the value of an item but not be willing to share that information despite pledges of confidentiality. Others may not know the value of an item that has a well-defined value, and they may not be willing or able to ascertain that value.[27] In other cases, there may be fundamental uncertainty about a value until the item faces a market transaction—for example, the value of a house or small business may only be knowable by soliciting bids.

---

[27]In the ideal, respondents would use all records at their disposal to provide as complete and accurate information as possible. In practice, interviewers must balance pressing respondents to use records against the possibility of alienating the respondent in what is already a burdensome interview. In the 1998 SCF, 6.8 percent of respondents referred to records frequently, but 66.0 percent of respondents never referred to them at all; the most commonly used records for those who used them were account statements, business records, and pension documents.

As far back in SCF history as the 1983 survey, interviewers were given a "range card" containing a series of dollar ranges with a letter associated with each range, and they were asked to present this card for respondents who might be able to provide partial information about the value. Coding conventions were developed to denote other types of bounding information that interviewers recorded, for example, "more than a million dollars." Although such bounding information was apparently not a negligible contribution, the approach was not as widely used by interviewers as one might expect. Part of the explanation for this sparseness of use may be that interviewers were traditionally taught to probe for and record data in such a way that would obscure the fact that there had been initial reluctance or indecision on the part of the respondent; for example, a respondent might have initially responded "between $500 and $1,000," the interviewer might have parried "would that be closer to $500 or to $1,000," and if the respondent gave either of the end points, the interviewer would have recorded that value. Another factor may be that there is only a weak incentive for interviewers to press respondents for information that would either slow down the interview or raise the possibility of alienating a reluctant respondent.

When the SCF migrated to CAPI in 1995, a decision was made to collect range data in a systematic way and to enforce a uniform minimum level of probing across all cases. A complex routine, "DKDOL," was constructed to allow great flexibility in collecting various types of partial information and to automate a large part of what would have been discretionary probing the past.[28] At the same time, the routine allows an interviewer to break out of the probing if the respondent becomes too resistant. Depending on the answers given by the respondent, the program records a single dollar value, a respondent-provided upper and lower bound, a response from a range card, or a range that results from leading the respondent through a decision tree. In every case, the program presents the interviewer with a confirmation screen that writes out in words the single amount or the endpoints of a range or partial range.

---

[28]An extensive description and analysis of the effects of this routine are given in Kennickell (1997).

Table 2: Reporting rates in percent for various items, combined AP and list samples, 1995 and 1998 SCF, unweighted.

Item	Have item		Value reported by respondent, for those reporting having the item					
	Yes	Unknown	Number	Range response			DK	Other missing*
				Tree	Card	$ range		
1995 SCF								
Credit card balance	76.0	0.4	93.6	0.4	4.1	0.2	0.1	1.7
Principal residence	67.6	0.0	88.9	1.1	7.2	1.1	0.0	1.7
Borrowed on mortgage	42.9	0.3	89.6	1.5	5.9	0.2	0.3	2.6
Owe on mortgage	42.9	0.3	86.1	1.6	8.5	0.1	0.2	3.5
Mortgage payment	42.2	0.3	92.7	0.4	4.2	0.0	0.1	2.5
Rent	23.8	0.0	95.1	0.4	2.9	1.0	0.0	1.5
Other real estate	32.4	0.6	84.0	1.7	9.5	0.7	0.4	3.7
Business	26.8	0.4	61.9	5.7	18.6	1.0	1.2	11.5
Car loan payment	23.7	0.2	93.0	0.8	4.1	0.0	0.2	1.9
Checking account	88.7	0.3	80.1	1.9	10.4	0.5	0.4	6.7
Money market account	17.3	0.7	71.7	1.8	14.4	0.5	0.9	10.6
Savings account	33.6	0.7	80.2	1.7	11.1	0.1	0.1	6.8
Certificates of deposit	17.0	1.0	69.7	3.4	11.1	0.3	0.3	15.3
IRA/Keogh account	34.6	1.2	74.4	2.6	13.5	0.3	0.4	8.9
Savings bonds	24.0	0.7	76.1	3.2	13.0	0.2	0.8	6.8
Municipal bonds	8.1	1.2	59.8	2.9	15.2	0.9	1.2	20.1
Tax-free mutual funds	8.3	1.6	59.6	2.5	16.6	0.0	0.8	20.5
Stock	28.4	0.9	63.8	2.5	16.0	2.2	1.4	14.1
Face value of whole life ins.	38.6	2.2	76.7	2.5	11.1	0.3	0.8	8.6
Cash value of whole life ins.	38.6	2.2	55.5	7.8	15.5	0.5	2.1	18.7
Wage income	73.6	1.0	72.8	1.5	16.7	0.2	0.3	8.4
Business income	20.6	1.5	68.5	2.4	12.8	0.3	0.5	15.6
Pension and Soc. Sec. inc.	26.5	1.2	73.3	1.9	11.1	0.0	0.4	13.3
Total income	100.0	0.0	69.1	1.5	16.8	0.1	0.5	12.1
1998 SCF								
Credit card balance	74.8	0.4	93.1	0.7	4.3	0.2	0.1	1.7
Principal residence	66.0	0.1	87.7	1.5	7.2	1.0	0.1	2.4
Borrowed on mortgage	41.7	0.4	86.3	1.8	6.8	0.2	0.2	4.6
Owe on mortgage	41.7	0.4	84.5	1.8	8.1	0.2	0.2	5.2
Mortgage payment	41.2	0.4	89.8	0.6	4.7	0.0	0.2	4.8
Rent	27.0	0.0	94.9	0.8	2.0	0.2	0.1	2.1
Other real estate	31.4	0.5	83.9	1.8	8.2	1.0	0.1	4.9
Business	26.6	0.4	65.4	6.4	13.2	1.6	0.5	12.8
Car loan payment	22.6	0.3	91.2	0.5	4.9	0.2	0.0	3.2
Checking account	89.0	0.5	77.0	2.2	10.1	0.9	0.3	9.6
Money market account	12.2	1.1	75.2	1.5	10.7	0.8	0.0	11.8
Savings account	50.6	1.1	75.1	2.3	11.2	0.5	0.1	10.8
Certificates of deposit	15.7	1.5	65.2	4.3	11.2	0.9	0.6	16.8
IRA/Keogh account	35.3	1.3	70.5	3.5	11.8	0.9	0.1	13.2
Savings bonds	19.8	1.3	74.0	3.4	10.4	0.2	0.4	11.6
Municipal bonds	6.8	2.0	63.7	2.7	13.0	1.0	0.0	19.5
Tax-free mutual funds	6.7	2.9	55.5	4.5	14.5	1.4	0.0	24.1
Stock	31.6	1.8	63.0	4.5	14.2	2.4	0.3	15.6
Face value of whole life ins.	36.2	2.7	74.9	3.0	8.1	0.3	0.8	13.0
Cash value of whole life ins.	36.2	2.7	54.2	8.4	11.6	0.4	0.9	24.4
Wage income	73.7	2.1	73.2	2.4	14.2	0.9	0.3	10.0
Business income	21.4	2.6	69.3	2.0	11.7	1.0	0.1	15.9
Pension and Soc. Sec. inc.	24.6	2.1	75.2	2.8	7.7	0.9	0.1	13.2
Total income	100.0	0.0	69.0	2.2	13.7	1.5	0.2	13.3

* Almost entirely refusals. Also includes a relatively small fraction of valuesset to missing in th course of data editing.

Several changes in the data were notable after the introduction of the DKDOL routine. The reporting of ranges increased substantially in frequency, while the frequency of complete responses declined somewhat and the frequency of the "don't know" response fell to the point that it became nearly negligible for most questions. At the same time, the frequency of final refusals did not change much, and direction of the changes was not consistent overall. These results suggest that the routine had the effect of forcing many interviewers to push for range data when they might not have done so in the past, and allowing some respondents who required probing to reach a single reported value in the past to give an explicit range response.

As shown in table 2, most of the range responses are ones given from the range card. Part of this use is explained by the question sequence in DKDOL, which asks first if the respondent can provide a range from the card. However, the program also allows the interviewer to enter a range more directly without having to go through this question, and a substantial fraction of the range card responses have this origin. It may be that the "threat" of having to go through the DKDOL sequence makes both interviewers and respondents more likely to minimize difficulties by going directly to the range card when the respondent does not give a complete response.

It is difficult to characterize how the resistance of respondents to answering questions may have changed over time because of the introduction of DKDOL. However, there does appear to be some trend since the introduction of that program in 1995. Final refusals at least on the dollar questions in table 2 have risen slightly in almost every instance (last column of the table).

Beginning with the 1989 SCF, missing data in the survey have been imputed using the FRITZ (Federal Reserve Imputation Technique Zeta) system, a multiple imputation (MI) model developed for the survey.[29] Each missing value in the survey is imputed five times, and the resulting data are stored as replicates of each observation, which are referred to as "implicates" to distinguish them from other data replicates.

---

[29]See Kennickell (1991, 1998) for a detailed discussion and evaluation of the multiple imputation in the SCF.

There are two great statistical advantages to MI: first, it is more efficient in that one can expect to get a more efficient estimate from multiple estimates of a missing value than from a single estimate—at least if there is any randomization involved in the imputation process, as there is in the SCF model. Second, MI makes clearer the uncertainty induced by having to make estimates based on partial information.

The FRITZ system is a sequential model, in that it follows a pre-defined sequence to impute every variable with missing data in the survey, and it is iterative, in that it multiply imputes all the missing data a number of times using the results of earlier imputations as a key input.[30] The model performs imputations of three types of variables: continuous, binary, and polychotomous variables. To give a basic sense of what the model does, the discussion here concentrates on the continuous variable case. The binary case is essentially the linear probability model extreme of the continuous variable model, and the polychotomous case is essentially a type of randomized hot deck procedure using conditional frequency tables.

At its core, the continuous variable imputations are based on a calculation that is roughly equivalent to a random draw from the predicted distribution of outcomes for a case based on a regression. The most interesting deviation from a straightforward regression is in the allowance for conditioning variables that may be missing. In theory, one should include all possible information that could determine the distribution of the missing data (see Little and Raghunathan (1997)), but degrees of freedom are too limited for such an approach except in surveys with far more observations than the SCF. In the FRITZ system, one specifies a "maximal" set of variables that might be used in the imputation. In the first iteration of the model, a covariance matrix is computed using all available pairs of variables in the maximal set. In later iterations, the covariance matrix is computed using the imputed data from the preceding iteration. Given an observation with a missing value for a variable at a particular point in the sequence of imputations, the model examines the values of the maximal set of conditioning variables for that observation and determines which values are not currently missing—that is, values that were

---

[30]Comparison of key estimates between iterations suggests that the system converges relatively quickly. In 1998, the process was stopped after six iterations. Of course, it may be the case that the model has only reached a temporary stationary point or that estimates of other factors would indicate less stability.

either originally reported by the respondent or have already been imputed within the iteration. These non-missing values are used to condition the model for the observation. FRITZ automatically subsets the rows and columns of the full covariance matrix to extract the relationships necessary for the more restricted regression. The imputation is made drawing randomly from the conditional distribution of the variable (where only the constant term, or equivalently the error term, is allowed to vary) until a draw is found that satisfies any prior constraints on the imputation. Such constraints may take many forms. For example, in some cases the imputation is subject to the bounds of a range reported by the respondent, and in other cases, the imputation may be constrained by institutional relationships or by logical relationships across variables. When an imputation is made, functionally related variables that are needed for conditioning further imputations or for other purposes are also computed.

A practical advantage of MI is that it makes it possible to analyze the data using standard statistical software that would be appropriate in the case that no data were originally missing with only a simple adjustment. However, the corresponding disadvantage for some purposes is that it makes it very difficult for analysts to ignore the fact of imputation. In many other surveys, there are no flags to indicate imputations, and even where there are such flags, it is not uncommon for researchers to ignore that information in analysis. Although it may be inconvenient at first to deal with the multiple imputations, it is actually straightforward. A standard error estimate that includes an adjustment for MI is given by

$$\sigma_{MI} = \{\overline{\sigma}_P^2 + [\sum_{i=1}^{5}(P_i - \overline{P})^2 * (1 + \frac{1}{n})/(n-1)]\}^{1/2}$$

where P is an estimated parameter (e.g., a regression coefficient, a mean, a median, a frequency count, etc.), $\overline{\sigma}_P^2$ is the average estimate across implicates of the variance of the estimate of P (under whatever model is used to estimate P), and n is the number of imputations (five in the case of the SCF, as noted earlier).

Although the FRITZ system works fairly dependably and the results of data simulation exercises suggest that the system is unbiased, there is still ample room for improvement. First, unlike more recent multiple imputation models, particularly those described in Schafer (1997) and Raghunathan *et al.* (1999), the FRITZ system limits randomization in the continuous

variable case to only a shift term rather than the theoretically more desirable alternative of drawing from the full posterior distribution of the data. The practical difference between such approaches is unknown at present. Some experimentation in this direction would be useful. Second, additional theoretical work needs to be done to support the type of sequential imputation used in the SCF. Currently, there is no theorem guaranteeing the existence of the distribution that FRITZ attempts to estimate. Third, it may be desirable to restructure the linear probability model used for the binary imputations to better approximate more theoretically desirable models such as a discriminant model or a logit. Finally, designing a system to automate variable selection and robustness checking would add enormously to the speed with which the imputations could be accomplished. The key problem here lies in the necessity of choosing a maximal set of variables that covers not just obvious direct correlates of the variable being imputed, but also variables that might serve as correlates of those variables should they also be missing. For observations with relatively few missing values, the model may have "too many" variables that are nearly collinear, and the model may become unstable.[31] A model selection procedure might limit the number of input variables and minimize the influence of poorly identified parameters.

3. Quality control and other data issues

Setting and maintaining standards of data quality require repeated attention to process design, training, measurement, evaluation, correction, documentation, and feedback. Because the world changes continually, this process never reaches a stationary state. Because much of such effort for the SCF has been discussed above already, this section highlights only a few additional areas of particular importance to the survey.

One activity that is pursued more intensely in the SCF than in most other surveys is the review and resolution of interviewers' comments. Interviewers are trained to record comments during the course of the interview when they encounter problems. In addition, for each completed case, the interviewers are required to answer a set of debriefing questions that ask about possible misclassification, reporting errors, and miscellaneous other factors that may bear

---

[31]Problems from exact collinearity are avoided by using a sweep algorithm for matrix inversion.

on the quality of the data recorded in the main instrument. At the same time, software developed over the course of problem detection and resolution in many earlier surveys is used to flag other potential problems. Past experience indicates that the resolution of the interviewers' comments and all the possible exceptions identified more mechanically makes a very substantial difference to the quality of the ultimate data.

In instances where the evidence is sufficiently clear that the respondent's data as recorded differ from what is correct, the data are changed by the project staff with subject matter expertise who perform the data review. Quite often, this work requires quite sophisticated inferences that would be far beyond a reasonable expectation of the performance of a "standard" survey editor. Since the introduction of CAPI, the data review has proceeded in tandem with the data collection. Thus, it is often possible to feed back information to the field staff in order to minimize future instances of problems. To prepare for the next wave of the survey, systematic problems with questions and procedures are cataloged both through written commentary and the error detection software.[32]

During the course of an interview, if a respondent's answer does not fit a category provided in the instrument or the interviewer is uncertain about the classification, the interviewer records a verbatim transcription of the response. This text information is reviewed by coders for possible resolution into responses identified in previous surveys as meaningful responses that are not sufficiently common to include directly in the instrument. Not infrequently, the coders are unable to resolve the response into a code. Standard practice in many surveys would be to treat all the remaining instances as "other" responses. In the SCF, such responses are reviewed again by project staff with subject matter expertise. Under current procedures, there are several possible outcomes at this stage: the response is resolved into an existing code, a new code is created, or the response is set to missing.[33] A response may be set to missing if the information

---

[32]Kennickell (1999a) provides an analysis of the errors identified in the editing of the 1998 SCF and provides a model of measurement error in terms of respondent and interviewer characteristics.

[33]See Kennickell (1999a) for more details. Prior to the introduction of CAPI, interviewers did not always record verbatim answers where appropriate; a decision was made in those surveys to allow unspecified "other" responses to remain in the final dataset. However,

provided did not answer the question asked or it was overly ambiguous. Because the creation of a new code is the least common outcome, the SCF experience suggests that some surveys that do not perform such intensive review may be misrepresenting the true frequency of "miscellaneous" events.

Every variable in the SCF has a parallel "shadow" variable that contains information about the content of the main variable. In the most straightforward instances, the shadow variable indicates that the data value is the original value provided by the respondent or that the question is not applicable to the respondent's circumstances. For dollar variables, the shadow variable also summarizes any range information provided by the respondent. A variety of other codes are used to record data changes. Although it is a practical impossibility to include sufficient information in one variable to allow a data user to reverse every change made after the initial data collection, these shadow variables do at least identify local changes and usually allow one to infer the nature of broader changes.

Throughout the history of the survey, efforts have been made to compare the survey findings with those in other sources. By focusing on areas where one might expect a relatively high level of conceptual agreement, one can get an indication of how effectively the SCF is measuring households' finances. Avery, Elliehausen and Kennickell (1988) compared 1983 SCF data with similar survey data in the Current Population Survey, the Survey of Income and Program Participation, and with aggregate data in the Federal Reserve Board flow of funds accounts. Johnson and Woodburn (1994) compared SCF data with estimates of wealth holdings derived from estate tax filings. Aizcorbe and Starr-McCluer (1997) looked at vehicle holdings in the SCF and the Consumer Expenditure Survey and compared those results with industry data. Antoniewicz (1996) has made a detailed comparison of the surveys from 1989, 1992 and 1995 with flow of funds data. Generally, the results of these comparisons are encouraging. Nonetheless, there are important discrepancies, and work to resolve them should continue.

---

[33](...continued)
with the enforced recording of verbatim information in CAPI, it is very rare for a code to be left unclassified; the instances of "other" codes in the final dataset are cases where the respondents' answers were meaningful, but so unusual as to be extremely unlikely to appear again in a later survey.

C. Disclosure limitation

The most perfectly realized survey would still be of little use if the information it collected were not made available to researchers. The SCF faces particularly great obstacles in meeting this goal.[34] The survey collects information that is generally viewed as quite personal, and it interviews a sample that includes a large number of people who are likely to be prominent at least in their local area. In addition, the wide variety of characteristics collected in the survey provide a view of households so detailed that virtually every case is unique in the sample, and it is likely that at least some cases are unique in the population.

In addition to constraints imposed by the Privacy Act and other such regulations, the survey operates under a set of contracts between the Federal Reserve Board and SOI, between the Federal Reserve Board and the survey contractor, and between the survey contractor and SOI; this structure is largely a result of requirements imposed in exchange for being allowed to use SOI data for the selection of the SCF list sample. The contracts specify that procedures for release of the SCF data must meet the strictest standards of any of the three parties. The effect of the agreement is that the release of SCF data must meet a test similar to that applied for the release of research files of tax data.

To limit disclosure, the SCF employs many of the same types of adjustments that are common in other surveys. In creating the public use dataset, some detailed codes are collapsed, dollar values are rounded, and some variables–including most geographic data–are suppressed. Some variables are top- or bottom-coded, although the use of such adjustments is generally for non-dollar variables. For a selection of cases, the limited geographic data released are manipulated in a way to undermine the usefulness of the information to a data intruder while minimizing the damage for legitimate research.

In a significant departure from other large surveys, the SCF also employs a data simulation technique for the public use dataset. For observations judged to be unusual in various ways and for a randomly selected set of additional observations, dollar values are multiply

---

[34]Fries, Johnson, and Woodburn (1996, 1997) provide a more detailed summary of the disclosure review for the survey along with a summary of key difference between the public and internal versions of the data. Kennickell (1998b, 1998d) describes and evaluates the data simulation technique use to limit disclosure in the public version of the SCF dataset.

simulated using the same models that are used in the multiple imputation of missing data. The simulation is performed in such a way that distortions of the covariation between variables and of key statistics like the mean are minimized. From relatively modest initial use of this technique in the 1989 SCF, the practice has progressed to use for every dollar variable in the set of cases to which it was applied in the 1998 survey. As a final measure, a set of unspecified and highly miscellaneous adjustments are made to the data to undermine the ultimate credibility of variables as keys to identifying a respondent; these adjustments are made in such a way that it is highly unlikely that they would have more than a negligible effect on legitimate research using the data.

Most of the information that can be provided to the public is available on the SCF website at http://www.federalreserve.gov/pubs/oss/oss2/scfindex.html in the form of codebooks and related technical documentation, research papers, survey instruments, and data.[35] There are two purposes for making this material readily available: First, users need to make informed choices in their analysis of the data; to this end, it would be convenient to include an engine that could be used to search the site for specific items. Second, it is hoped that by making the data creation process as open as possible, others who are less close to the project may be able to see areas that are in need of improvements and share their insights with the project staff.

III. Directions for Future Research

Virtually all surveys present serious challenges to data collectors: An appropriate questionnaire must be developed, a suitable sample of respondents must be selected, interviewers must be trained and motivated, the respondents must be persuaded to participate, they must be encouraged to provide complete and accurate data, the data must be processed consistently, and the data must be made available to the public In creating wealth data, all these challenges are amplified. This paper has focused on the measurement of wealth in the Survey of Consumer Finances with the goal of providing a comprehensive, if relatively abbreviated, view

---

[35]Relatively voluminous and specialized materials such as interviewer training manuals are not included on the site, but copies are available upon request.

of the methodological foundations of the SCF as they stand now, and to point to the directions where further development is needed.

The history of wealth measurement in the SCF is a good example of evolution. In the 1983 survey, the agendas of the original sponsors were quite different and no one involved knew what the survey would ultimately turn into. What allowed the project to define its place more clearly over time has been a simplification of the sponsorship and, more importantly, a healthy research atmosphere where it is reasonable to recognize both strengths and flaws with clarity and to feed that understanding forward to subsequent designs. Such feedback is central to the survey. For example, a key high frequency process involves timely reactions to interviewers' electronic comments during the field period, and key lower frequency processes are holding debriefings and keeping logs of easily forgotten events to be used in the next iteration of the survey. Given the proven importance of feedback in the evolution of the project, perhaps even that process should be reexamined for areas needing improvement.

Several specific points in the survey where further work is needed were highlighted in this paper. The most pressing sample design issue is achieving a less noisy stratification of the list sample; for this purpose the most promising avenue is using multiple years of SOI data, and efforts toward this end are currently underway. Because for some purposes the sample size of the SCF is too small to allow sufficiently precise estimates, it would be desirable to expand the number of observations in both the list and AP samples; however, given the large cost of such a step, it would be hard to justify at this time.

Unit nonresponse analysis is very likely to continue to be of the highest level of importance to the quality of the survey results. In addition to developing new tools for the field staff to use in persuading respondents and continuing research on nonresponse, two other steps should be pursued: First, attention should be paid to the incentives operating on interviewers and to the likely consequences of such incentives on the distribution of the types of cases ultimately interviewed; research in this area is ongoing. Second, all surveys should recognize that the issue of unit nonresponse is a common one and pool some of their efforts. It may be productive to work through a neutral scientific organization, such as the American Statistical Association or the National Academy of Sciences, to develop a program for educating people about the value of survey research. If nothing acts in a continuing way to counter the steady criticisms of surveys

in the popular media, all surveys are likely to face even greater challenges to maintain response rates in the future.

In the area of instrument design several factors are important. First, identification and investigation of problem questions must continue. Second, it may be important to recognize more explicitly that there is a limit to the ability of question designers to find question language that is unambiguous to everyone; advances in CAPI may aid in reducing error, but there may also be a need to do more to model the variation in understanding of questions by both interviewers and respondents. Third, we have a considerable distance to go in improving screen designs for CAPI; it is widely recognized that good design of paper questionnaires is important in reducing error, and it is clear from many of the sorts of errors in the SCF CAPI data that with flexible screen design we could make it easier for interviewers to do the right thing and harder to do the wrong thing.

Although imputation for missing data in the SCF functions reasonably smoothly, research needs to continue in order to make the process both more robust and faster. Finally, disclosure concerns are likely to grow over time as private databases on individuals grow even larger; although not an ideal solution, the data simulation methodology in the SCF offers hope for continuing to protect the privacy of respondents, and work in this area should be continued while searching for more appealing longer-term solutions.

Bibliography

Aizcorbe, Ana and Martha Starr-McCluer, "Vehicle Ownership, Vehicle Acquisitions and the Growth of Auto Leasing," *Monthly Labor Review*, Vol. 120, No. 6 (June 1997), pp. 34-40.

Antoniewicz, Rochelle [1996] "A Comparison of the Household Sector from the Flow of Funds Accounts and the Survey of Consumer Finances," Finance and Economics Discussion Series 1996-26, Board of Governors of the Federal Reserve System (June).

Avery, Robert B. and Gregory E. Elliehausen [1985] "1983 Survey of Consumer Finances: Technical Manual and Codebook," http://www.federalreserve.gov/pubs/oss/oss2/83/scf83home.html.:

Avery, Robert B., Gregory E. Elliehausen, and Arthur B. Kennickell [1988] "Measuring Wealth with Survey Data: An Evaluation of the 1983 Survey of Consumer Finances," *Review of Income and Wealth* (December), pp. 339-369.

Curtin, Richard T. [1985] Survey of Consumer Finances: Employer Sponsored Pension Benefit Plans Richard T. Curtin [1985], University of Michigan

Durkin, Thomas A and Gregory E. Elliehausen [1978] *1977 Consumer Credit Survey*, Board of Governors of the Federal Reserve System, Washington, DC

Frankel, Martin and Arthur B. Kennickell [1995] "Toward the Development of an Optimal Stratification Paradigm for the Survey of Consumer Finances," paper presented at the 1995 Annual Meetings of the American Statistical Association, Orlando, FL.

Fries, Gerhard, Barry W. Johnson, and R. Louise Woodburn [1996] "Disclosure Review and its Implications for the 1992 Survey of Consumer Finances," paper for presentation at the 1996 Joint Statistical Meetings, Chicago, IL.

Fries, Gerhard, Barry W. Johnson, and R. Louise Woodburn [1997] "Analyzing Disclosure Review Procedures for the Survey of Consumer Finances," paper for presentation at the 1997 Joint Statistical Meetings, Anaheim, CA.

Groves, Robert M. and Mick P. and Couper [1996] "Contact-Level Influences in Face-to-Face Surveys," *Journal of Official Statistics*, Vol. 12, No. 1, pp. 63-83.

Heeringa, Steven G., Judith H. Conner and R. Louise Woodburn [1994] "The 1989 Surveys of Consumer Finances Sample Design and Weighting Documentation," working paper, Survey Research Center, University of Michigan, Ann Arbor, MI.

Internal Revenue Service [1992] "Individual Income Tax Returns, 1990."

Johnson, B. W. & Woodburn, R. L. (1994), "The Estate Multiplier Technique: Recent Improvements for 1989," Compendium of Federal Estate Tax Data and Personal Wealth Studies, Department of Treasury, IRS Publication 1773, pp. 391-400.

Katona, George, James N. Morgan, Hay Schmiedeskamp, and John A. Sonquist [1968] *1967 Survey of Consumer Finances*, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

Kennickell, Arthur B. [1991] "Imputation of the 1989 Survey of Consumer Finances: Stochastic Relaxation and Multiple Imputation," 1991 Proceedings of the Section on Survey Research Methods, Annual Meetings of the American Statistical Association, Atlanta, GA.

Kennickell, Arthur B. [1997] "Using Range Techniques with CAPI in the 1995 Survey of Consumer Finances," *Proceedings of the Section on Survey Research Methods*, 1996 Annual Meetings of the American Statistical Association, Chicago, IL.

Kennickell, Arthur B. [1998a] "Using Income Data to Predict Wealth," paper presented at the Annual Meetings of the Allied Social Science Associations, New York, 1999a.

Kennickell, Arthur B. [1998b] "Multiple Imputation and Disclosure Protection: The Case of the 1995 Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B.[1998c] "List Sample Design for the 1998 Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B.[1998d] "Multiple Imputation in the Survey of Consumer Finances," *Proceedings of the Section on Business and Economic Statistics*, 1998 Annual Meetings of the American Statistical Association, Dallas, Texas.

Kennickell, Arthur B. [1999a] "Measuring Data Quality in the 1998 Survey of Consumer Finances," *Proceedings of the Section on Survey Research Methods*, 1999 Annual Meetings of the American Statistical Association, Baltimore, MD.

Kennickell, Arthur B. [1999b] "Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances,"*Journal of Official Statistics*, v. 15 no. 2, 1999, pp. 283-304.

Kennickell, Arthur B.[1999c] "How Different Are the "Late" Cases? Evidence from the 1998 Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. [1999d] "Revisions to the SCF Weighting Methodology: Accounting for Race/Ethnicity and Homeownership," working paper, Board of Governors of the Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. [2000] "Codebook for the 1998 Survey of Consumer Finances," http://www.federalreserve.gov/pubs/oss/oss2/98/scf98home.html.

Kennickell, Arthur B. [2000b] "An Examination of Changes in the Distribution of Wealth From 1989 to 1998: Evidence from the Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. [2000c] "Asymmetric Information, Interviewer Behavior, and Unit Nonresponse," working paper, Board of Governors of the Federal Reserve Board. http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. and Douglas A. McManus [1993] "Sampling for Household Financial Characteristics Using Frame Information on Past Income," *Proceedings of the Section on Survey Research Methods*, 1993 Annual Meetings of the American Statistical Association, San Francisco, CA.

Kennickell, Arthur B, Douglas A. McManus, and R. Louise Woodburn [1996] "Weighting Design for the 1992 Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board. http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. and Janice Shack-Marquez [1992] "Changes in Family Finances from 1983 to 1989: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, (January) pp. 1-18.

Kennickell, Arthur B. and Martha Starr-McCluer [1997a] "Household Saving and Portfolio Change: Evidence from the 1983-89 SCF Panel," *Review of Income and Wealth* (Series 43, No. 4), December, pp. 381-399.

Kennickell, Arthur B., Martha Starr-McCluer, and Annika E. Sundén [1997] "Household Saving and Financial Planning: Some Findings from a Focus Group," *Financial Counseling and Planning*, Volume 8(1), pp. 1-17.

Kennickell, Arthur B. and Martha Starr-McCluer [1997b] "Retrospective Reporting of Household Wealth: Evidence from the 1983-89 Survey of Consumer Finances," *Journal of Economic and Business Statistics,* October.

Kennickell, Arthur B, Martha Starr-McCluer and Brian J. Surette [2000] "Recent Changes in Family Finances: Results from the 1998 Survey of Consumer Finances," *Federal Reserve Bulletin*, (January 2000) pp. 1-29.

Kennickell, Arthur B. and R. Louise Woodburn [1992] "Estimation of Household Net Worth Using Model-Based and Design-Based Weights: Evidence from the 1989 Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board. http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. and R. Louise Woodburn [1999] "Consistent Weight Design for the 1989, 1992, and 1995 SCFs, and the Distribution of Wealth," *Review of Income and Wealth* (Series 45, number 2), June, pp. 193-215.

Little, Roderick J.A. and Trivellore Raghunathan [1997] "Should Imputation of Missing Data Condition on All Observed Variables?" *Proceedings of the Section on Survey Research Methods*, 1997 Joint Statistical Meetings, Anaheim, California.

Projector, Dorothy S. and Gertrude S. Weiss [1966] "Survey of Financial Characteristics of Consumers," Board of Governors of the Federal Reserve System.

Raghunatha, Trevellore E., P.W. Solenberger, P.W., and J.van Hoewyk [1999] "IVEWARE: Imputation and Variance Estimation Software," Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor.

Schafer, Joseph L. [1997] *Analysis of Incomplete Multivariate Data*, London: Chapman and Hall.

Tourangeau, Roger, Robert A. Johnson, Jiahe Qian, Hee-Choon Shin, and Martin R. Frankel [1993] "Selection of NORC's 1990 National Sample," working paper, National Opinion Research Center at the University of Chicago, Chicago, IL.

Wilson, O., and Smith, W.J. Jr. (1983) "Access to Tax Records for Statistical Purposes," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 595-601.